Wavelet Approximation in Data Assimilation

Andrew Tangborn
Data Assimilation Office NASA-GSFC, Code 910.3, Greenbelt, MD
and
JCET, University of Maryland-Baltimore County, Baltimore, MD

Abstract

Estimation of the state of the atmosphere with the Kalman filter remains a distant goal because of high computational cost of evolving the error covariance for both linear and non-linear systems. Wavelet approximation is presented here as a possible solution that efficiently compresses both global and local covariance information. We demonstrate the compression characteristics on the the error correlation field from a global two-dimensional chemical constituent assimilation, and implement an adaptive wavelet approximation scheme on the assimilation of the one-dimensional Burger's equation. In the former problem, we show that 99% of the error correlation can be represented by just 3% of the wavelet coefficients, with good representation of localized features. In the Burger's equation assimilation, the discrete linearized equations (tangent linear model) and analysis covariance are projected onto a wavelet basis and truncated to just 6% of the coefficients. A nearly optimal forecast is achieved and we show that errors due to truncation of the dynamics are no greater than the errors due to covariance truncation.

Introduction

Data Assimilation is in general the estimation of the state of a system achieved through the combination of observations and a physical/dynamical model of the system. Most data assimilation systems use Bayesian estimation theory to obtain an optimal estimate of the state. Achieving this optimal estimate requires the minimization of a *cost function* which in turn requires a statistical knowledge of the forecast and observation errors. In schemes where the error statistics are not evolved in time, such as statistical interpolation or 3DVAR, errors are generally assumed homogeneous and isotropic, inspite of the numerous physical sources of inhomogeneity (*eg* fronts), Desroziers (1993)). .In the Kalman filter, the forecast errors consist of propagated initial errors and errors created by the model (which are in turn propagated forward). Errors statistics are also altered by the observation network, which itself is highly non-uniform. Thus the propagated error fields produced by the Kalman filter contain the inhomogeneities which represent both the physical and data driven variations in accuracy. Localized features like fronts are resolved to the extent that the computational grid allows. However, propagation of forecast errors is perhaps the single most computationally expensive component of a data assimilation system (though rapid increases in observations may change this). Approximation of the propagation step therefore has become an important area of investigation.

Numerous techniques which approximate the forecast errors and their propagation have been proposed and tested. Approximation means that some part of the error covariance must be neglected and ideally one should neglect the part that has the least impact on the assimilation results. These include evolution of error variance only (Dee, 1991), SVD and eigen-decomposition (Todling and Cohn, 1994), (Tippet *et al* 2000), Error subspace estimation (Lermusiux and Robinson, 1999) and wavelet representation (Chin *et al.* 1994,1999), (Tangborn and Zhang, 2000). Farrell and Ioannou (2001a,b) have developed a "balanced truncation" method which retains both leading terms in both the covariance and the dynamics in terms of Hankel singular values. The ensemble Kalman filter (Evenson, 1994; Houtekamer and Mitchell, 1998) evolves error statistics using an ensemble of forecasts rather than applying the dynamical model (or its tangent linear model) to the full error covariance.

In present work, we revisit the wavelet approximation scheme in order to understand

how to understand how both non-linearity and the longer term evolution will affect the efficiency of wavelet expansions to represent and propagate error covariances. We focus in particular on the scale features of error covariances and how they influence the success or lack of success of the wavelet approximation on different dynamical systems. A major concern is the adaptivity of the scheme, so that there is a means to include wavelet coefficients that have been truncated, but become important in later stages of the assimilation. The issue of truncation of the propagator dynamics raised by Farrell and Ioannou is also addressed.

The major difference between wavelet representation and the other schemes is local representation. Atmospheric dynamics are inherently local, and it is local structures which often drive larger scale phenomena. For example crossing a front might require less than two or three grid points, and there is no guarantee that the most energetic EOFs or singular values will capture it. Yet a front may impact atmospheric motion over much larger scale regions. Error covariances need to be able to resolve these frontal structures because they are regions of large error variance. A higher weighting to observations will result only if the error covariances are correctly calculated on these scales.

Wavelet representation of data allows for this localization, and therefore has the potential to overcome some of these obstacles to efficient storage and propagation of error covariances. Wavelet coefficients may in fact be a useful way to uncover localized information on model error. In this paper we explore the representation and propagation of forecast errors by wavelet expansion. Our approach is to determine not only when the scheme is successful, but when it fails (as the number of retained coefficients are reduced). In this way we can determine what features in the error covariance are essential to achieving optimal or near optimal assimilation.

The particular wavelet transform used in this work comes from the family of compactly supported orthonormal wavelets introduced by Daubechies (1988). One could proceed by employing a wavelet-Galerkin scheme (Restrepo and Leaf, 1995) for solving the covariance evolution equation, but this presents a number of difficulties. These can be bypassed by starting with discretized equations from another numerical scheme and projecting them onto a wavelet basis (Tangborn and Zhang, 2000). This approach is particularly attractive because of the many existing data assimilation systems that could be adapted to these approximations.

The wavelet transform results in coefficients that are local in location and scale, and at each finer scale, the wavelength reduces by half. At each scale there then are twice as many coefficients. We use the standard notation of j to represent scale and k to represent translations. The total number of scales represented by a transform of n points is $J = \log(n)/\log(2)$ and the number of translations in a given scale j is $K = 2^j$. Because each coefficient represents a given scale *and* location (unlike spectral representation in which only scale is represented), there is a greater flexibility to retain important localized features. This is particularly important when only a fraction of the coefficients are retained.

In this paper we display the wavelet coefficient representation of covariances (and correlations) as a single column for each dimension. Images of wavelet coefficients will vary from coarse to finer scales (as with other spectral representation), but the finest scale is half of the coefficient vector.

In the first part of the paper we examine the evolution of the forecast correlation from Kalman filter experiments carried out by Menard *et al.* (2000) and Menard and Chang (2000). The propagation of error correlations for chemical constituents in the stratosphere is governed essentially by the horizontal wind field. The resulting correlation structure becomes highly complex with a great deal of small scale structure. In the second part of this paper we demonstrate the wavelet approximation of the error covariance propagation step in the extended Kalman filter applied to the 1-d Burger's equation.

Representation of Constituent Correlation Errors

Figure 1 shows the evolution of the error correlation from the Kalman filter assimilation of CLAES CH_4 over a 7 day period. An initial first order autoregressive (FOAR) correlation model is used as the initial condition (a), and is given by

$$\mathbf{C}(i,j) = exp\Big(-\frac{|\mathbf{r}_i - \mathbf{r}_j|}{L}\Big). \tag{1}$$

where $\mathbf{r}_i$ and $\mathbf{r}_j$ are the position vectors of to points on the sphere. After several days of wind driven movement the correlations becomes highly dependent on the local wind fields, with a wide range of significant length scales, as shown in Figure 1.

When we project these correlations onto a wavelet basis (Daub12), we find that 32 (out of 2048 coefficients) contain 99.7 % of the total energy of the initial correlation function. On days 3,5 and 7, the energy contained in the largest 32 coefficients is 97.7, 99.3 and 99.2 % respectively. Thus, even as the complexity of the correlation field grows, there is no significant decrease in the efficiency of the wavelet representation in terms of the energy contained in the first 32 coefficients. We define the fraction of energy in a reconstruction as

$$\frac{E}{E_{tot}} = \Big(\frac{\sum_{k=1}^{n_{trunc}} \hat{C}_k^{\,2}}{\sum_{k=1}^{n_{total}} \hat{C}_k^{\,2}}\Big)^{1/2} \tag{2}$$

where $\hat{C}$ is the k^{th} largest wavelet coefficient of the correlation function C(i,j). Figures 2(a-b) show the evolution of the wavelet representation of the correlation at the same times. In each direction (longitude and latitude), the wavelet coefficients are ordered from coarser to finer scales. At scale J there are 2^J translations and the finest scale will contain half the coefficients. From Figure 2(a), we see that initially just few coefficients contain most of the energy of the correlation field. There are no significant small scale coefficients. At later times (Fig. 2(b-d)) there is a small increase in the magnitudes of a few of the smaller scale coefficients, but only in the meridional direction. The stretching in the zonal direction tends to create short scale variations primarily in the zonal meridional direction. Thus, there are essentially no small scale wavelet coefficients in the zonal direction, even after 7 days. This type of structure keeps the dimensionality of the wavelet representation at a minimum since only a tiny fraction of the coefficients have any significant magnitude.

Reconstruction of the correlations from the 32 largest coefficients (via fast wavelet inverse transform) results in the correlations shown in Figures 3(a-d). A qualitative inspection shows

that much of the localized structure in the original correlations has been preserved in the low dimensional wavelet representation. The reconstructions are of course not perfect, and the question arises as to what has been lost, and what what scale. This can be determined by taking the difference between Figures 2 and 3, transforming the result to wavelet space and plotting the result for each scale separately. Figure 4(a) shows the difference between the full and 32 term approximation of the correlation on day 7.

The maximum error due to the wavelet approximation is about 0.1, at any given scale. If we project these differences onto the wavelet basis and separate out the scales (by saving only the j_{th} scale in each case) before applying the inverse transform we can identify which scales contribute the most to the truncation errors. The two-dimensional wavelet transform, like the 2-d Fourier transform, allows for length scales in different directions to be specified independently. Since the grid is 64×32, there are 6 (j=0-5) scales in the zonal direction and 5 scales (0-4) in the meridional direction. Not all of these 30 possible scale combinations are important (ie contain larger errors), and we summarize them by plotting the maximum error for each combination in Figure 4(b). In the longitudinal direction, when the latitudinal scales are large, the maximum errors occur at $j_x = 2$, which corresponds to an average scale of 90 degrees. At finer latitude scales, the maximum error in the zonal direction is also concentrated in the finer scales. The dependency on latitudinal scale is an almost monotonically increasing truncation error with decreasing scale. The largest absolute error, occurs at $j_x = 3$ and $j_y = 3$, which corresponds to an average scales of 90 and 45 degrees in the zonal and meridional directions, respectively. The errors incurred in the correlation by truncating the wavelet expansion to 32 of 2048 total coefficients are never larger then 0.1, and at coarser scales are always less than 0.05.

Some examples of the scale decomposition of the truncation errors are shown in Figures 5(a-h). Figures 5(a-d) are all at scale j=3 (22.5^o) in the meridional direction and vary from j=2 (90^o) to j=4 (22.5^o) in the zonal direction. As the zonal scale decreases, the truncation errors become larger and are concentrated at the edge of the non-zero correlation region. A similar pattern is seen in the j=4 (11.25^o) meridional scale (Figures 5(e-h)). This error analysis suggests that a significant portion of small scale correlation (and therefore covariance) structure can be represented when just 32 out of 2048 (or 1.6 %) wavelet coefficients are retained. Correlation loss due to wavelet truncation occurs along the edge where correlation values are relatively small. It doesn't appear however, that information on the *location* of

this edge has been lost (compare Figures 1(d) and 3(d), for example. This may indicate that the structure essential to successful data assimilation is included in this small subspace.

Propagation of Error Covariances in Wavelet Space

The successful representation of a variety of correlation scales is only part of the goal of the wavelet approximation presented here. A much harder problem is the propagation of error covariances by evolving only the subspace. This approach opens up a host of difficulties, including changes to the propagator structure and evolution of the covariance subspace itself. We first deal with the issue of structural changes to the propagator. The propagator $\mathbf{\Psi}$ is really just a representation of the matrix system,

$$\mathbf{A}\mathbf{u}^{k+1} = \mathbf{B}\mathbf{u}^{k} \tag{3}$$

by

$$\mathbf{u}^{n+1} = \mathbf{\Psi}\mathbf{u}^{n} \tag{4}$$

where $\mathbf{u}^k$ is the state of the system at time t_k and $\mathbf{\Psi} = \mathbf{A}^{-1}\mathbf{B}$. In most NWP models, $\mathbf{A}$ has some banded structure, which is taken advantage of when choosing a solution algorithm. The actual computation is based on 3 while 4 is a convenient notation for presenting the propagation portion of a data assimilation system. The evolution of the forecast error covariance $\mathbf{P}^f = \langle(\mathbf{u}^t - \mathbf{u}^f)(\mathbf{u}^t - \mathbf{u}^f)\rangle$ can then be calculated by

$$\mathbf{P}^f_{k+1} = \mathbf{\Psi}\mathbf{P}^f_k\mathbf{\Psi}^T \tag{5}$$

The numerical algorithm is actually carrying out

$$\mathbf{A}\mathbf{P}^f_{k+1}\mathbf{A}^T = \mathbf{B}\mathbf{P}^a_k\mathbf{B}^T \tag{6}$$

The analysis error covariance is projected onto a wavelet basis

$$\hat{\mathbf{P}} = \mathbf{W}\mathbf{P}\mathbf{W}^T \tag{7}$$

where $\mathbf{W}$ is the matrix representation of the fast wavelet transform. In order to carry out the covariance propagation, we need to project the propagator onto the wavelet basis as well,

$$\hat{\mathbf{A}} = \mathbf{W}\mathbf{A}\mathbf{W}^T \tag{8}$$

$$\hat{\mathbf{A}}^T = \mathbf{W}\mathbf{A}^T\mathbf{W}^T \tag{9}$$

$$\hat{\mathbf{B}}^T = \mathbf{W}\mathbf{B}\mathbf{W}^T \tag{10}$$

$$\hat{\mathbf{B}}^T = \mathbf{W}\mathbf{B}^T\mathbf{W}^T \tag{11}$$

Structural changes to the matrices **A**,**B** depend on their initial form. One might be concerned that the sparse banded structure that occurs in finite difference discretization of convection-diffusion type problems might be destroyed by the transform to wavelet space. Quite fortunately, the reverse occurs: matrices become more sparse. Ground breaking work by Beylkin *et al.* (1991) showed that diagonally dominant matrices become sparse when projected onto a wavelet basis. This is another sort of compression since fewer terms are needed to represent the operator, thereby reducing the operation count. The meaning of this *near* diagonalization of the propagator is that there is much less interaction between the wavelet coefficients for the covariance during than there is in physical space. This is a natural effect of the orthogonality and localization of the basis.

We can demonstrate this on a relatively simple system, namely the finite difference approximation to the 2-d convection diffusion equation (univariate). This system, when discretized as $n \times n$, results in a sparse pentadiagonal matrix structure ($n^2 \times n^2$) where the band width is $2n$. Solution of this system generally requires $O(n^3)$ operations. Figure 6 shows a 40×40 pentadiagonal matrix (a) and its projection onto the Daubechies wavelet basis having 12 filter coefficients (b)(henceforth called Daub12). In wavelet space the matrix is no longer pentadiagonal, having a still sparse *hierarchically band diagonal* form. If we can truncate the very small off diagonal terms (they are about $1/100^{th}$ the size of the diagonal terms, we have then ended up with a diagonal system. Other options include using solvers created specifically for solving systems of this form (Beylkin *et al*, 1991) which require only $O(N)$ operations. The end result is a system, already reduced by a factor of 50 or more which can be solved by an $O(N)$ algorithm.

In the previous section we demonstrated the compression characteristics of the wavelet transform on error correlations from constituent assimilation by *reordering* the wavelet coefficients by descending magnitude before truncating. We must therefore address the impact of coefficient reordering on the matrix structure. If the error covariance is simply reordered according to the magnitude of each element, we cannot retain the essentially diagonal form

of Figure 6(b). However, if the reordering is done only for the purpose truncation, and the retained coefficients (and the corresponding matrix elements) are kept in their original order, then this structure will also be retained.

The ordering and subsequent truncation of the wavelet coefficients also introduces the problem of loss of potentially important dynamics from the propagator when the TLM. Farrell and Ioannou (2001a,b) dealt with this issue by using a balancing transformation to obtain a representation in coordinates where the stochastic optimals (fastest growing modes) and the empirical orthogonal functions (EOFs) coincide. The simple system of figure 6 is normal and no special treatment is needed. However, the TLM of Burger's equation in the next section is non-normal and dynamics truncation may result in some eventual covariance loss. While there is not yet a general method for ensuring that the important parts of the dynamics are retained, we can make some conclusions about this on a case by case basis.

Another issue that needs to be addressed is adaptivity of the scheme. As the error covariance evolves in time, we expect that the wavelet coefficients or *modes* that are largest should change. For example, as a frontal system moves in time, it should carry with it the associated larger error variances. Since the wavelet representation is local, coefficients in the region that the front moves into should grow and become more important. How can we deal with resulting variability in the coefficient ordering?

This problem is partially resolved by the assimilation step itself. We must transform the propagated error covariance back to physical space in order to calculate the gain matrix for weighting the observations and forecast. At the same time, the analysis covariance is calculated in physical space using both the gain matrix and the propagated error covariance. Therefore the analysis covariance will depend on the observation network. When this new analysis covariance is projected onto the wavelet basis, a new set of coefficients will be reordered and truncated. Modes that may have been previously truncated may at this time be retained, particularly if the observation network varies in time. This insures that changes in the covariance structure due to the observation network do impact the ordering of coefficients. But this will only partially solves the issue of adaptivity since we haven't really taken into account movement taking place during the propagation step. This problem can be seen in Figure 7 which shows the evolution of the magnitudes of the 32 largest coefficients (a) and the smallest 1948 coefficients (b) over the 7 day period. Some of the coefficients in

(a) grow in time, but a number decay dramatically, as advection alters the correlation field. At the same time, some of the small coefficients in (b) increase enough to become larger than many of the the initially large coefficients. Thus the initial ordering of coefficients will change significantly over the 7 day simulation period.

When we carry out a the propagation of error correlations, we don't know *a priori* which coefficients are going to become important in the future. We only know which ones are large for the initial conditions. If one makes the simple (but incorrect) assumption that the coefficient ordering won't change significantly in time, then a loss of representation should be found. We tested this by using the coefficient ordering, with truncation at 32 terms, from the initial correlation in Figure 1(a). On days 3,5 and 7 the fraction of energy retained is now 92.0 %, 96.9 % and 97.2 %, (down from 97.7,99.3,99.2) respectively. This may appear to be a minor decrease, but the shorter length scales contain smaller energy, as defined by 2. The loss in representation should occur in the smaller scales since the initial correlation has no small scale structures. We plot the correlation reconstruction from day 7 using these 32 term expansions in Figure 8 which show a significant loss in small scale details of the correlation, as expected.

We have now shown that the wavelet transform applied to the propagation equation for error covariances can result in both reduction of the system size and an increase in the sparcity of the system matrix. However, if the list of retained wavelet coefficients is not constantly updated, the approximation will become poor in time, as the winds move finer scale features to parts of the domain. If we can develop a scheme that does retain the largest coefficients, then it is possible to vastly reduce the computational cost of error covariance propagation in data assimilation without significant loss in accuracy. In the next section we demonstrate this method on a simple non-linear system using the extended Kalman filter (EKF).

The Extended Kalman Filter (EKF)

We have chosen the Burger's equation to demonstrate the wavelet approximation scheme for several reasons. Solutions to this non-linear equation tend to contain sharp frontal features (particularly when small viscosity is used). Localized structure of this short is generally difficult to approximate by traditional expansions (*ie.* spectral, eigendecomposition, SVD).

We wish to compare the success of the assimilation scheme with both the full extended Kalman filter and other approximation schemes. Burger's equation with stochastic forcing is used as the dynamical core of the Kalman filtering system.

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} = \nu\frac{\partial^2 u}{\partial x^2} \tag{12}$$

with periodic boundary conditions:

$$u(0,t) = u(1,t)$$

and initial conditions:

$$u(x,0) = f(x)$$

Discretization is carried out using second order centered difference in space and Adams-Bashforth in time. The resulting system in matrix form is then

$$\mathbf{A}\mathbf{u}^{n+1} = \mathbf{B}\mathbf{u}^n \tag{13}$$

The matrices $\mathbf{A}, \mathbf{B}$ are state dependent because of non-linearity. The tangent linear model (TLM) of the non-linear dynamical system is created by perturbing the initial condition at one grid point and carrying out the computation for one timestep, for each grid point. This generates an estimate of the Jacobian matrix from which the tangent linear model propagator $\mathbf{\Psi}^{TLM}$ can be created. The forecast is updated by the non-linear system:

$$\mathbf{u}_f^{n+1} = \mathbf{\Psi}\mathbf{u}_f^n \tag{14}$$

An added noise vector $\mathbf{w}_m^n$ represents the model error and is Gaussian distributed, with correlation length L_c.

$$\mathbf{u}_t^{n+1} = \mathbf{\Psi}\mathbf{u}_t^n + \mathbf{w}_m^n$$

where $\mathbf{u}_f^{n+1}$ is the forecast at time t^{n+1}, and $\mathbf{u}_t^{n+1}$ is the true state. $\mathbf{\Psi} = \mathbf{A}^{-1}\mathbf{B}$ is the propagator. An observation process is defined by

$$\mathbf{u}_o = \mathbf{H}\mathbf{u}_t + \mathbf{w}_o \tag{15}$$

where where $\mathbf{u}_o$ is the observed value, $\mathbf{H}$ is the observation matrix, and $\mathbf{w}_o$ is the observational error, which is also Gaussian distributed.

The full extended Kalman filter consists of the following steps:

update forecast:

$$\mathbf{u}_f^{n+m} = \mathbf{\Psi}_m \mathbf{u}_a^n \tag{16}$$

$\mathbf{\Psi}_m = \mathbf{\Psi}^m$ represents m timesteps of the discretized system equations. Update true state (as a substitute for a real physical process to observe)

$$\mathbf{u}_t^{n+m} = \mathbf{\Psi}_m \mathbf{u}_t^n + \sum_{i=n}^{n+m} \mathbf{w}^i \tag{17}$$

Update forecast covariance

$$\mathbf{P}_f^{n+m} = \mathbf{\Psi}_m^{TLM} \mathbf{P}_a^n \mathbf{\Psi}_m^T + \mathbf{Q}_m \tag{18}$$

Calculate Kalman gain:

$$\mathbf{K}^{n+m} = \mathbf{P}_f^{n+m} \mathbf{H}^T (\mathbf{H} \mathbf{P}_f^{n+m} \mathbf{H}^T + \mathbf{R}^{n+m})^{-1} \tag{19}$$

Update analysis variable

$$\mathbf{u}_a^{n+m} = \mathbf{u}_f^{n+m} + \mathbf{K}^{n+m} (\mathbf{u}_o^{n+m} - \mathbf{H} \mathbf{u}_f^{n+m}) \tag{20}$$

Update analysis covariance

$$\mathbf{P}_a^{n+m} = (\mathbf{I} - \mathbf{K}^{n+m} \mathbf{H}) \mathbf{P}_f^{n+m} \tag{21}$$

where $\mathbf{P}_f$ is the forecast covariance matrix, $\mathbf{P}_a$ is the analysis covariance matrix, $\mathbf{Q}$ is the model error covariance and $\mathbf{R}$ is the observational error covariance. $\mathbf{K}$ is the Kalman gain matrix.

The operation count for the covariance propagation step (18) is $O(N^3)$ if we do not take advantage of the banded structure of the matrices, (where N is the number of grid points). If the number of observations at each analysis time is significantly less than N, then (18) is the most computationally intensive step in the assimilation system. It is this step on which we are focusing the present work. The next section outlines a scheme for approximating the error covariance propagation by a truncated wavelet expansion.

An Adaptive Wavelet Approximation to the EKF System

The wavelet-truncation scheme involves projecting both the tangent linear propagator $\mathbf{\Psi}^{TLM}$ and the analysis covariance $\mathbf{P}_a$ onto the wavelet basis and only a small fraction of the coefficients. The covariance propagation is then carried out in wavelet space on the smaller system.

We repeat the Kalman filter algorithm here with an adaptive Kalman filter approximation scheme.

Calculation of the Kalman gain and updating of the state variable $\mathbf{u}$ are carried out in physical space.

$$\mathbf{u}_f^{n+m} = \mathbf{\Psi}_m \mathbf{u}_a^n \tag{22}$$

$$\mathbf{u}_t^{n+m} = \mathbf{\Psi}_m \mathbf{u}_t^n + \sum_{i=n}^{n+m} \mathbf{w}^i \tag{23}$$

Project analysis covariance onto wavelet basis:

$$\hat{\mathbf{P}}_a^n = \mathbf{W}^T \mathbf{P}_a^n \mathbf{W}$$

Project covariance propagator onto wavelet basis:

$$\hat{\mathbf{\Psi}}^{TLM} = \mathbf{W}^T \mathbf{\Psi} \mathbf{W}$$

The important question here is how to decide which part of $\hat{\mathbf{\Psi}}$ and $\hat{\mathbf{P}}_a^n$ to retain. Simply truncating after L terms is equivalent to removing all the smallest scales. We showed in the constituent assimilation example that smaller scales can be more important than larger scales. In this Burger's equation example we create small scale features in one dimension particularly as the viscosity is decreased. In a sharp frontal system, forecast errors tend to be larger and very localized because any velocity (or dispersion) errors are amplified. It is also important that the resulting forecast covariance is positive definite so that the Kalman gain equation can be solved. Any truncation of the wavelet space representation of $\hat{\mathbf{P}}_a$ is carried out on entire rows and columns, rather than on individual matrix entries.

The matrix system in equation 16 is projected onto a wavelet basis as described by equations (6-11)

$$\hat{\mathbf{P}}_f^{n+m} = (\hat{\mathbf{\Psi}}_m)(\hat{\mathbf{P}}_a^n)(\hat{\mathbf{\Psi}}_m^T) + (\hat{\mathbf{Q}}_m) \tag{24}$$

In order to retain the most important components of the covariance the rows and columns of $\hat{\mathbf{P}}_\mathbf{a}$ are re-ordered using the magnitude of the main diagonal as the ordering criterion. Note that in wavelet space, this diagonal term contains information on both the variance and correlation fields. Simultaneously, we must reorder the covariance propagator $\hat{\mathbf{\Psi}}^{TLM}$ and its transpose as well.

After reordering, the covariance propagation equations are truncated to to L terms and covariance propagation is carried out in wavelet space.

$$(\hat{\mathbf{P}}_f^{n+m})_L = (\hat{\mathbf{\Psi}}_m)_L(\hat{\mathbf{P}}_a^n)_L(\hat{\mathbf{\Psi}}_m^T)_L + (\hat{\mathbf{Q}}_m)_L \tag{25}$$

Transform forecast covariance to physical space:

$$\mathbf{P}_f^{n+m} = \mathbf{W}\hat{\mathbf{P}}_f^{n+m}\mathbf{W}^T \tag{26}$$

Update Kalman gain:

$$\mathbf{K}^{n+m} = \mathbf{P}_f^{n+m}\mathbf{H}^T(\mathbf{H}\mathbf{P}_f^{n+m}\mathbf{H}^T + \mathbf{R}^{n+m})^{-1} \tag{27}$$

Calculate Analysis:

$$\mathbf{u}_a^{n+m} = \mathbf{u}_f^{n+m} + \mathbf{K}^{n+m}(\mathbf{u}_o^{n+m} - \mathbf{H}\mathbf{u}_f^{n+m}) \tag{28}$$

Update the analysis covariance

$$\mathbf{P}_a^{n+m} = (\mathbf{I} - \mathbf{K}^{n+m}\mathbf{H})\mathbf{P}_f^{n+m} \tag{29}$$

We have carried out an ensemble of 15 assimilations for each of 4 wavelet expansions of the covariances, L=4,8,16 and 128. Each assimilation differs in both the initial condition and model error perturbations, but all have the same statistics. We initially assume that both the initial, model error and observation statistics are known perfectly. The full (untruncated) and approximate EKF are compared for Burger's equation on the domain

$$0 \leq x \leq 1$$

using initial conditions on $\mathbf{u}$

$$u_o(x) = sin(2\pi x) x \leq 0.1 \tag{30}$$

$$u_o(x) = 0 \quad 0.1 \leq x \leq 1.0 \tag{31}$$

and run parameters $\delta t = 0.01$, $m = 20$, $U = 1$ and $\nu = 0.005$.

We have carried out an ensemble of 15 assimilations for each of 4 wavelet expansions of the covariances, L=4,8,16 and 128. Each assimilation differs in both the initial condition and model error perturbations, but all have the same statistics. We assume that both the initial, model error and observation statistics are known perfectly.

The model error and observational error covariances have correlation lengths of $L_c = 0.02$ and variance of 0.0001. The simulation are run for 180 timesteps with observations made every 20 timesteps. There are observations at every grid point for $0.38 \leq x \leq 1$ and none elsewhere.

The initial $u_o(x)$ field from equations 30 and 31 is shown in Figure 9 (a). Advection will result in the spike in $u(x)$ to move to the right, while viscosity will cause a decay in the peak. The non-linearity of Burger's equation maintains a steep front on the right side of the peak. Figure 9 (b) shows the ensemble average $\langle u^f \rangle$ at the final state for the full set of coefficients (L=128) and partial expansions (L=8,16). The initial spike has moved to the right, just into the observed region ($x \geq 0.38$). The only significant differences in the final state u_f between the different expansions is in the unobserved region.

The ensemble mean of the rms errors $(u - u_t)$ for of each case is shown in figure 10. A 87 % reduction in the number of coefficients (L=16) results in rms errors only marginally higher than the full Kalman system. A further reduction to L=8 (a 94 % reduction) only increases the error by about 5 %, whereas the L=4 case (97 % reduction) results in a roughly 15 % rise in the peak rms error.

We have also calculated the energy retained in the approximated covariance matrices, as a basis of comparison with the covariance approximation done on the constituent assimilation. The 4,8 and 16 term expansions represent 3, 6 and 12 % of the coefficients in one dimension (or about 0.1, .4 and 1.6 % if extended to two dimensions). The energy retained in each 1-d case is found (by equation 2) 54, 72 and 81 % respectively. Therefore, in the one-dimensional Burger's equation, we only need to retain about 74 % of the energy by retaining 6 % of the coefficients in order to achieve assimilation results close to the full system.

In order to understand how approximating the error covariance by truncating its wavelet expansion (equation 25) affects the assimilation results of Figures 9 and 10 we need to consider representation errors in the error covariance itself. We consider the variance and covariance fields separately. Figure 11 shows the diagonal of the forecast covariance matrix,$\mathbf{P}_f$ (an estimate of the error variance) at the end of the assimilation for the full system (bold line) and wavelet approximations which retain L=32 (thin line), L=16 (grey line) and L=4 (dash-dot) wavelet coefficients. All of the approximate error variances have some loss that decreases as the number of retained terms is increased. More significantly, both the L=8 and

L=16 at least partially retain the peak in error variance around $x = 0.35$ that is associated with the frontal structure in u. Because the L=16 case results in a very slight increase in the rms error, the accurate approximation of the variance spike in the observation void region appears to be a key component of the covariance approximation.

Regions of steep gradients tend to result in the largest forecast errors because small dispersive type errors can cause errors of large amplitude. When larger forecast errors are coupled with the substantial underestimate of $L = 4$ case in Figure 11, the Kalman filter will weight the forecast to much and result in the larger errors of Figure 10. In particular, the $L = 4$ case completely misses the error variance peak, indicating that the Kalman filter will be unable to distinguish between regions with and without steep gradients. The result is the

Figure 12 shows a plot of a row from $\mathbf{P_f}$ corresponding to the position $x = 0.35$ at the end of the assimilation. This plot shows the *covariance* between this position and other points within the system. Here we see that the $L = 8, 16$ wavelet expansions accurately approximate the covariance except at the peak, which corresponds to the variance shown in Figure 11. The coarsest approximation $L = 4$ again has significant degraded accuracy. Not only is most of the covariance structure lost, but the correlation length is significantly increased. Since the covariance is a product of both the variance and the correlation, this indicates that the forecast error correlation is accurate down to $L = 8$ and significant increases in error occur for smaller expansions.

We next consider whether the loss of accuracy is due to truncation of the covariance itself or to truncation of the propagator, equation . Figure 13 shows the difference between the full and truncated wavelet representation along the main diagonal ($\hat{\mathbf{P}}_\mathbf{f} - (\hat{\mathbf{P}}_\mathbf{f})_L$ for example) for the cases $L = 8, 16$. For $L = 16$, the difference in the diagonal are very small up to (and including) the 16^{th} and then become large. This jump at the 17^{th} coefficient is due entirely to the truncation of the covariance, and the propagator truncation has no impact. For the $L = 8$ case, we see that the differences in the first 8 coefficients are on the same order as the larger coefficients, indicating that propagation errors have become important. When the expansion is truncated at 4 coefficients, the propagator truncation errors become more important than covariance truncation errors, since the 4^{th} coefficient difference is the largest. Thus we see that only when the system becomes most severely truncated do the

errors incurred by truncating the dynamics Ψ dominate. This confirms the qualitative argument made about the reduction of significant modes in the projection of Figure 6(b).

We have also calculated the energy retained in the approximated covariance matrices, as a basis of comparison with the covariance approximation done on the constituent assimilation. The 4,8 and 16 term expansions represent 3, 6 and 12 % of the coefficients in one dimension (or about 0.1, .4 and 1.6 % if extended to two dimensions). The energy retained in each 1-d case is found (by equation 2) 54, 72 and 81 % respectively. Therefore, in the one-dimensional Burger's equation, we only need to retain about 74 % of the energy by retaining 6 % of the coefficients in order to achieve assimilation results close to the full system.

Finally, we return to the question of the impact of truncation of the TLM on the TLM. While the TLM used here for Burger's equation is non-normal, the coefficient ordering has been such that, except for the first eight terms, the coefficient magnitudes decrease monotonically. Thus, as long as the first few coarsest scales are retained (up to j=3 in this case), the remaining decay exponentially. In this TLM, wavelet representation results in a nearly diagonal matrix with entries that decay rapidly towards zero as the coefficient index increases, as shown in Figure 14. Thus truncation of finer scales past some cutoff will not result in a loss of the most significant dynamics in this particular case.

Conclusions

We have investigated the structure of the projection of the error covariance matrix onto a wavelet basis for two disparate problems: two-dimensional constituent assimilation and the one-dimensional Burger's equation. In in the former it is shown that accurate representation of the covariances (99 % of the energy) using a small fraction (1.5 %) of the coefficients. The largest errors in this approximation occur in regions of steepest gradients of the covariance (a kind of constituent front). In the Burger's equation assimilation we demonstrate that the representation is accurate enough to obtain nearly optimal assimilation results using an extended Kalman filter when about 6 % of the coefficients 74 % of the energy and are retained. The assimilation results in sharply higher rms errors when the number of retained coefficients is dropped below this level (3 % of coefficients). At this point the forecast error covariance matrix does not resolve the spike in variance occurring at the front and results in a correlation length that is too long. Significant loss of the dynamics is also seen in this level

of truncation. The important conclusion from this failure is that these important covariance characteristics can be attained with just 6 % of the coefficients resulting in a near optimal assimilation.

The need for adaptivity is demonstrated in the constituent assimilation example, while the Burger's equation assimilation shows how it can work. In spite of the relative simplicity of the one-dimensional example, it contains a number of the features of more complex systems. The adaptive scheme is able to capture the high variance region associated with the moving front, and severe truncation of the wavelet coefficients does not cause significant loss in dynamics. Extension of this scheme to multi-dimensional constituent assimilation is shown to be viable by these results and is goal of ongoing work.

References

Beylkin, B., Coifman, R. and Rokhlin, V. Fast Wavelet Transforms and Numerical Algorithms, *Communications on Pure and Applied Mathematics*, **41**, pp 141-183, 1991.

Chin, T.M. and A.J. Mariano. Wavelet-based Compression of Covariances in Kalman Filtering of Geophysical Flows, *Proc. SPIE*, **2242**, 1994.

Chin, T.M., A.J. Mariano and E.P. Chassignet, Spatial Regression and multiscale approximations for sequential data assimilation in ocean models, *J. Geophys. Res.* **104**, 7991-8014, 1999.

Daubechies, I., Orthonormal bases of Compactly Supported Wavelets. *Communications in Pure and Applied Mathematics*, **41**, 909-996, 1998.

Dee. D.P., Simplification of the Kalman Filter for Meteorological Data Assimilation, *Quart. J. Roy. Meteor. Soc.*, **117**, 365-384, 1991.

Desroziers, G. and J-P. Lafore, A Coordinate Transformation for Objective Frontal Analysis, *Mon. Wea. Rev.*, **121**, pp. 1531-1553, 1993.

Farrell, B.F. and P.J. Ioannou, Accurate Low-Dimensional Approximation of the Linear Dynamics of Fluid Flow, *Journal of the Atmospheric Science*, **58**, pp. 2771-2789, 2001.

Farrell, B.F. and P.J. Ioannou, State Estimation Using a Reduced-Order Kalman Filter, *Journal of the Atmospheric Sciences*, **58**, pp. 3666-3680, 2001.

Kalman, R.E., A New Approach to Linear Filter and Prediction Problems. *Trans . ASME, Ser., D, J. Basic Eng.*, **98**, pp. 35-45, 1960.

Houtekamer, P.L. and H.L. Mitchell, Data Assimilation Using an Ensemble Kalman Filter Technique, *Mon. Wea. Rev.*, **126**, pp 796-811, 1998.

Lermusiaux, P.F.J. and A.R. Robinson, Data Assimilation via Error Subspace Statistical Estimation. Part I: Theory and Schemes. *Mon. Wea. Rev.*, **127**, 1385-1407, 1999.

Lermusiaux, P.F.J., Data Assimilation via Error Subspace Statistical Estimation. *Mon. Wea. Rev.*, **127**, 1408-1432, 1999.

Menard, R., Cohn, S.E., Chang, L.-P. and Lyster P.M., Stratospheric Assimilation of Chemical Tracer Observations Using a Kalman Filter, Part I: Formulation. *Mon. Wea. Rev.*, **128**, 2654-2671, 2000.

Menard, R. and Chang, L.-P., Stratospheric Assimilation of Chemical Tracer Observations Using a Kalman Filter. Part II; χ^2 - Validated Results and Analysis of Variance and Correlation Dynamics. *Mon. Wea. Rev.*, **128**, 2672-2686.

Kalman, R.E., A New Approach to Linear Filter and Prediction Problems. *Trans . ASME, Ser., D, J. Basic Eng.*, **98**, pp. 35-45, 1960.

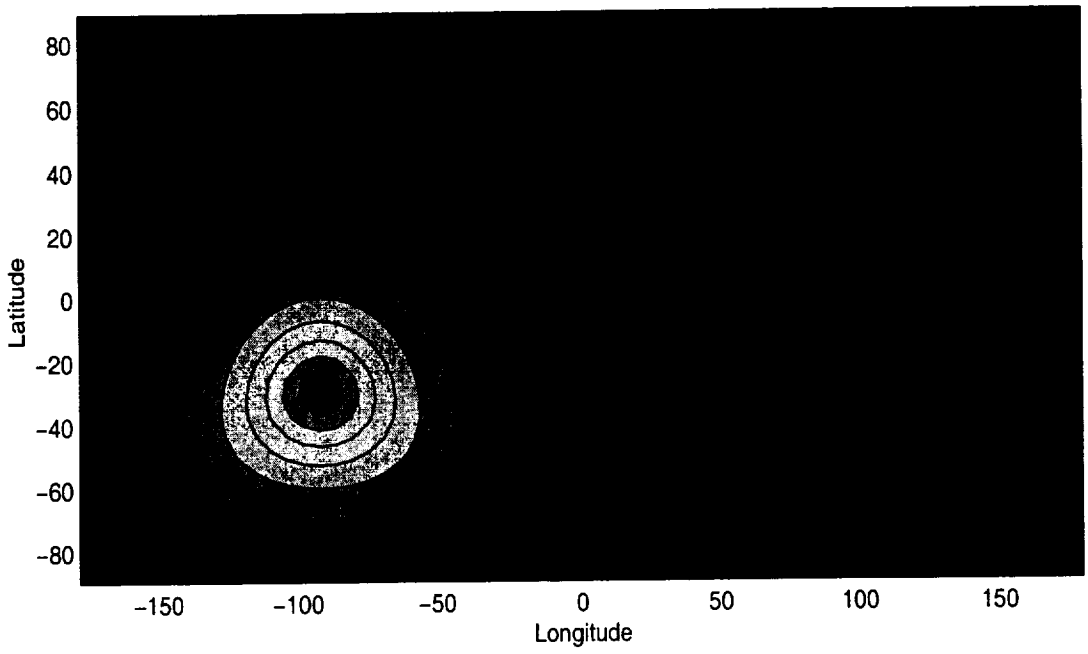
Restrepo, J. and G. Leaf., Wavelet-Galerkin Discretization of Hyperbolic Equations, *J. Comp. Physics*, **122**, 118-128, 1995.

Riishojgaard, L.P., A Direct Way of Specifying Flow-Dependent Background Error Correlations for Meteorological Analysis Systems, *Tellus*, **50A**, 43-57, 1998.

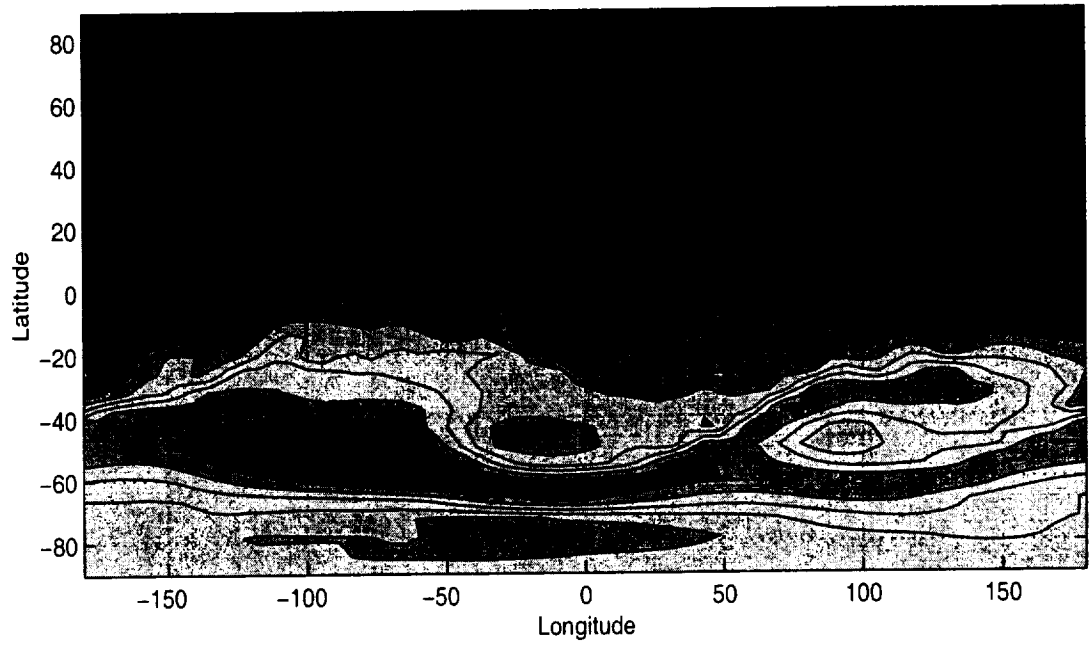
Tangborn, A. and Zhang, S, Wavelet Transform Adapted to an Approximate Kalman Filter System, *Applied Numerical Mathematics*, 2000.

Tippet, M.K., S.E. Cohn, R. Todling and D. Marchesin, Low-dimensional representation of error covariance, *Tellus A*, **52**, 533-553, 2000.

Todling, R. and S. Cohn, Suboptimal Schemes for Atmospheric Data Assimilation Based on the Kalman Filter, *Mon. Wea. Rev.*, **122**, pp 2530-2557, 1994.

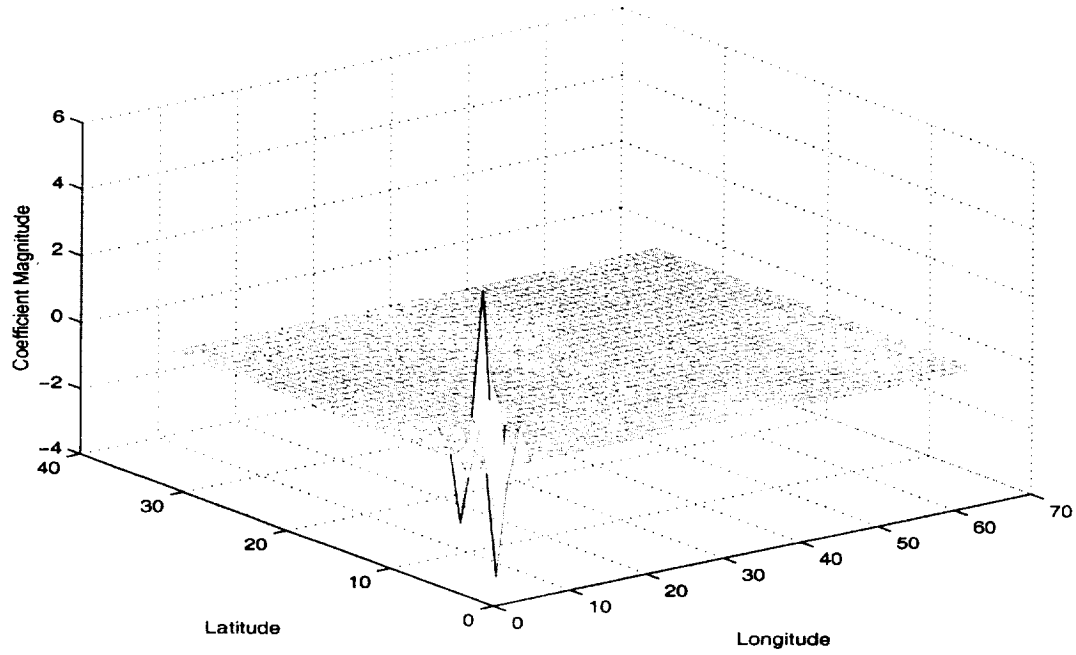


(a) Day 1

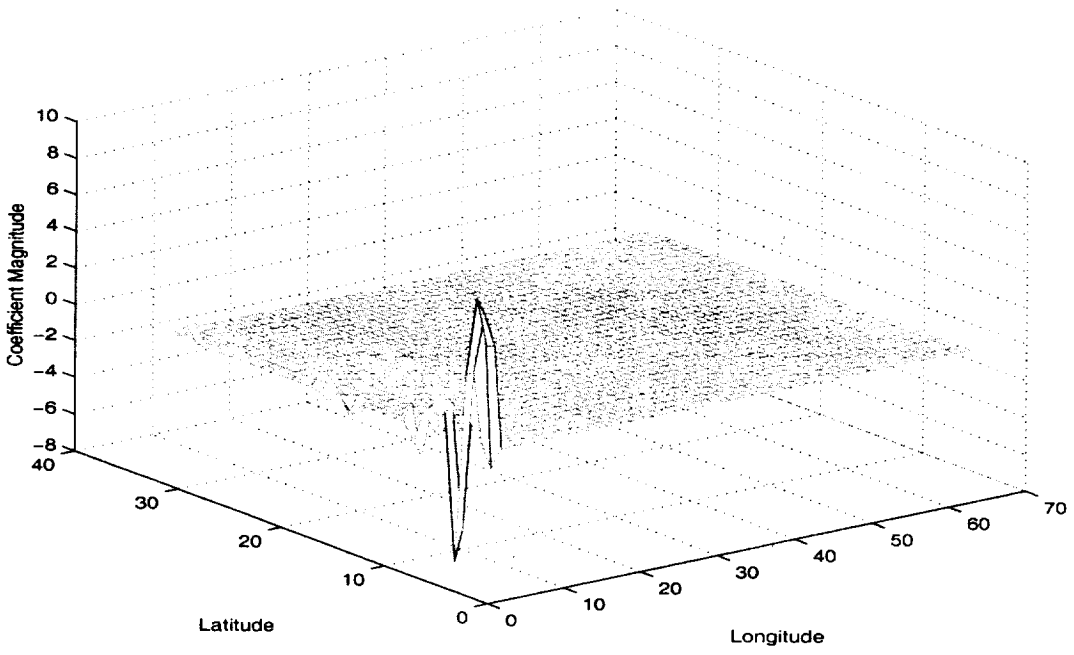


(b) Day 7

Figure 1: 7 day evolution of error correlation field for CH_4. The contour levels are equally space from 0 to 1 (0.5 is the lightest shade).

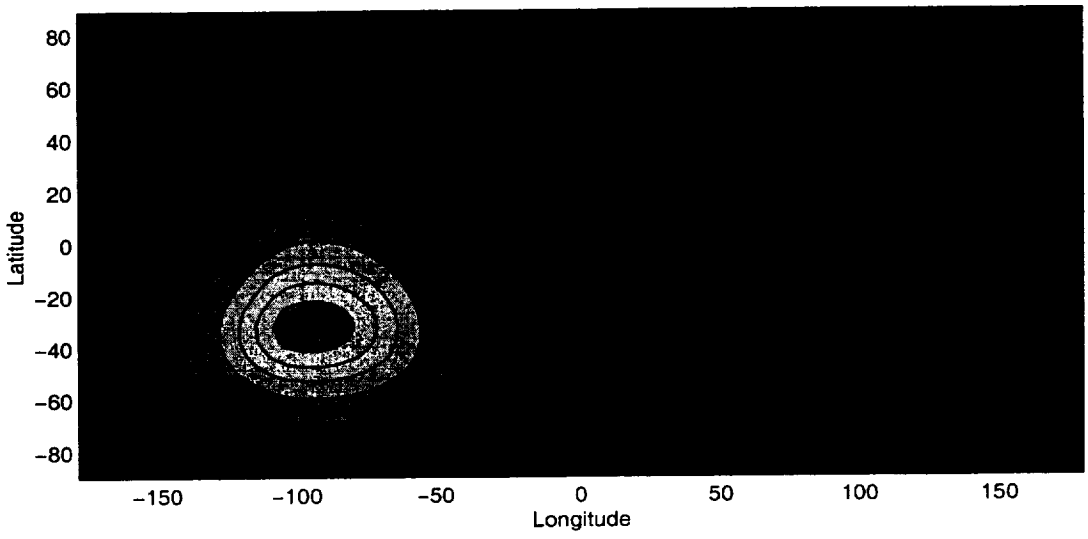


(a) Day 1

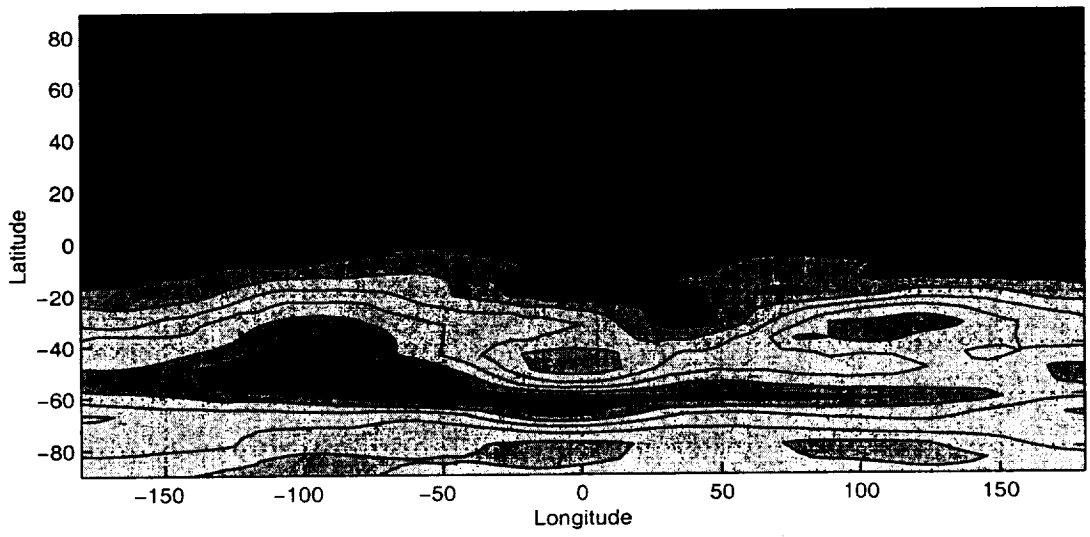


(b) Day 7

Figure 2: 7 day evolution of error correlation wavelet coefficients. Note that the move towards finer scales is much more pronounced in the latitude than longitude, due to primarily zonal direction shearing of the field.



(a) Day 1



(b) Day 7

Figure 3: 7 day evolution of the 32 term wavelet reconstruction of the error correlation field. 32 wavelet coefficients corresponds to $(1/64)^{th}$ of the coefficients. Growth in finer scales in the meridional direction can be seen on day 7.

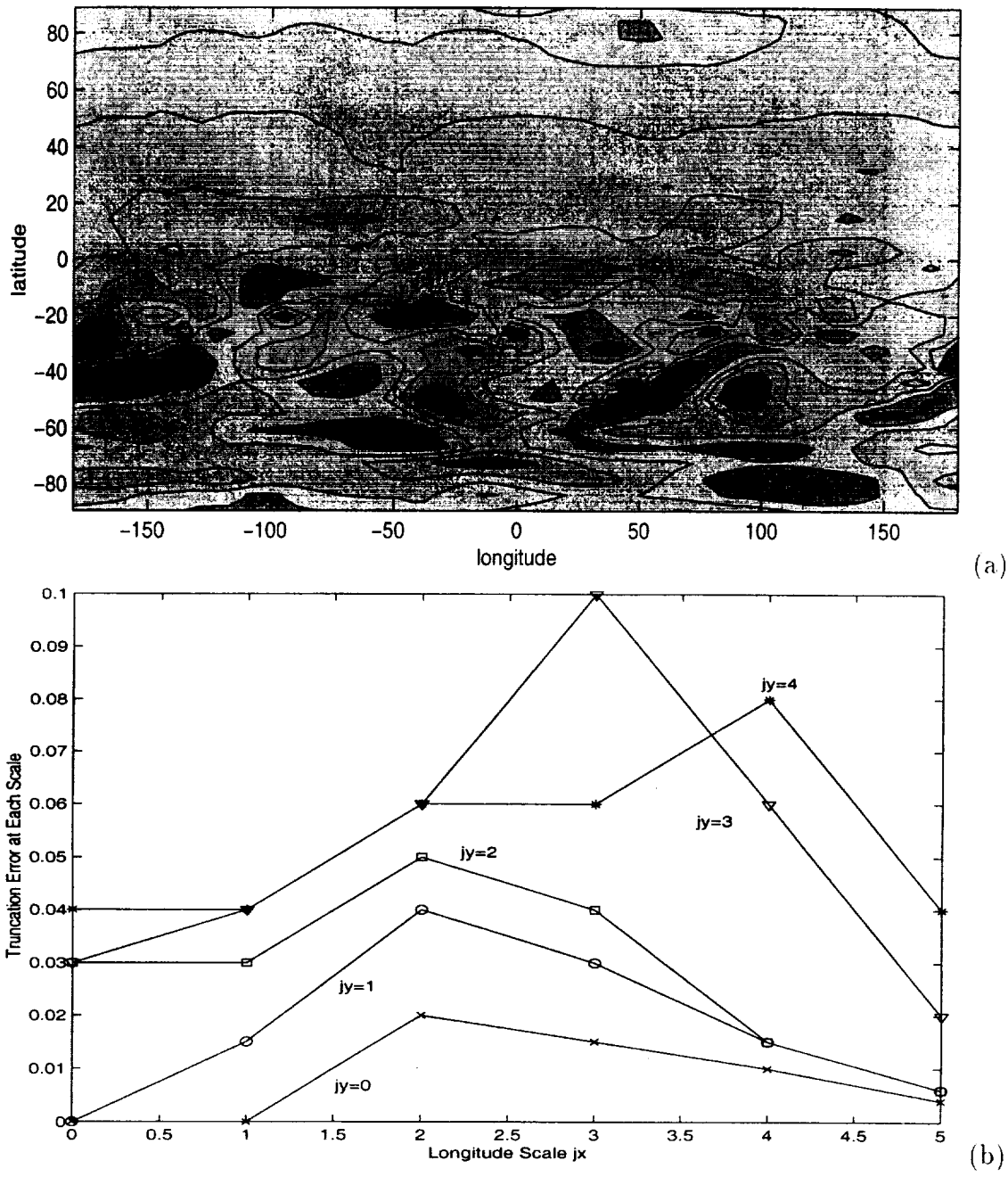
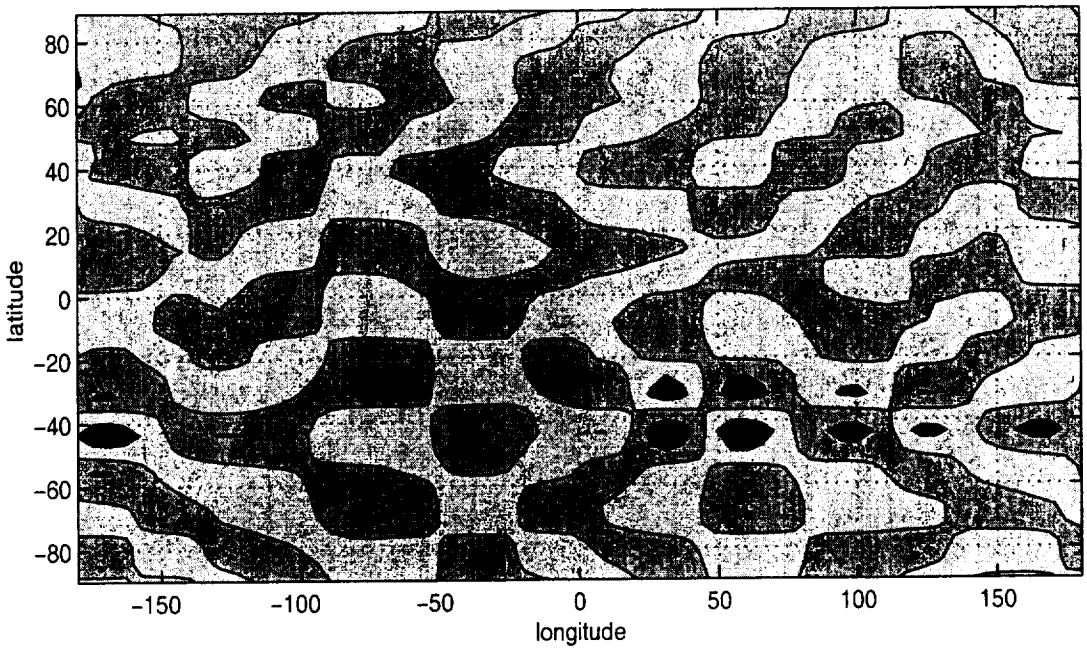
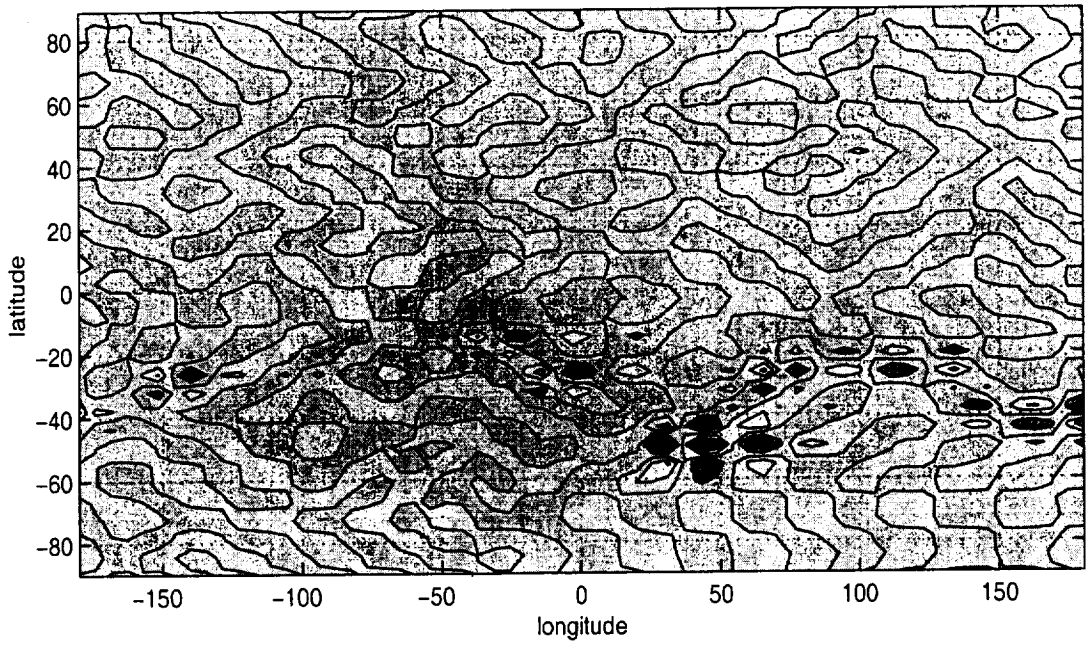


Figure 4: Difference between full correlation and 32 term wavelet approximation on day 7 (a) and the maximum truncation error for each scale combination (b)

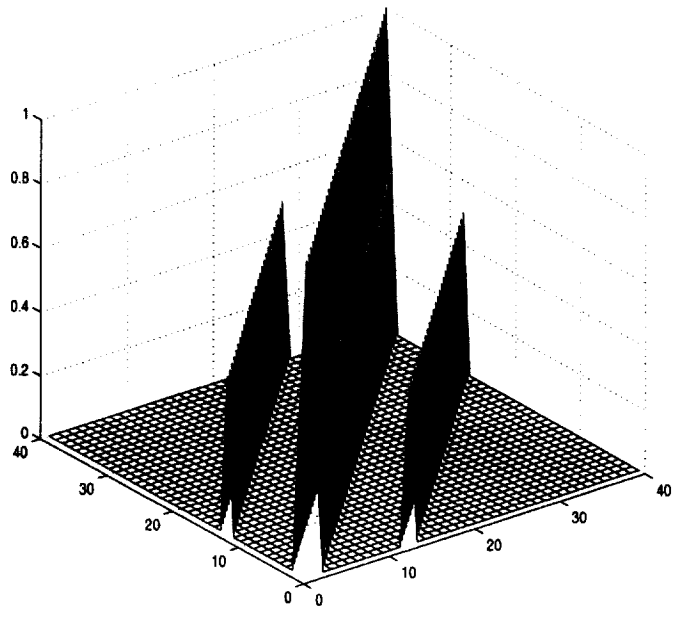


(a)

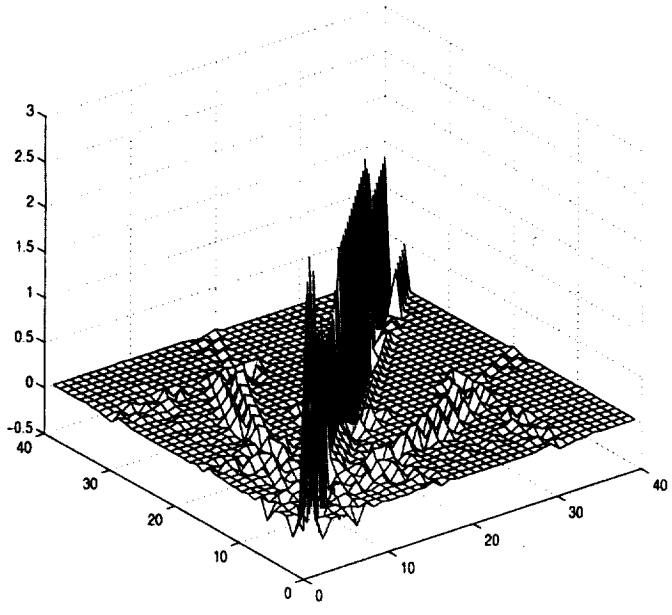


(b)

Figure 5: Differences between full and 32 term truncation of the error correlation on day 7, separated by scale. Shown are the two scales showing the maximum truncation error in the covariance, $j_x = j_y = 3$ (a) and $j_x = j_y = 4$ (b). These correspond to length scales of 22.5^o in latitude and 45^o in longitude for (a) and 11.25^o in latitude and 22.5^o in longitude for (b).



(a)



(b)

Figure 6: (a) 40x40 penta-diagonal matrix. Note that the two off diagonals are so close to the main diagonal that they appear as one. (b) Projection onto Daub12 wavelet basis. The finest scale is represented by half the indices in each direction.

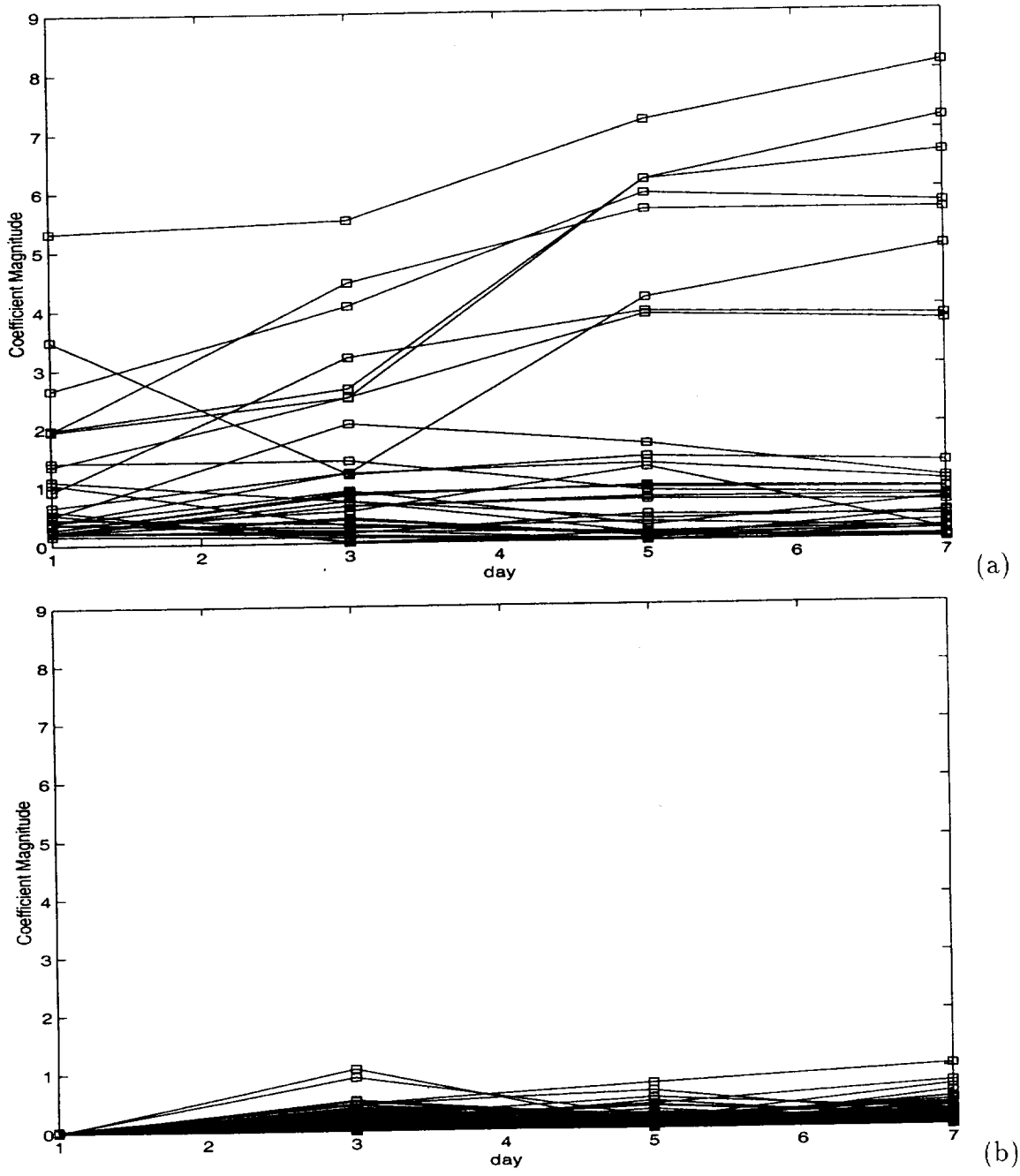


Figure 7: Evolution of the 32 initially largest (a) and the initially smallest 948 coefficients (b) wavelet coefficients for the error correlation for constituent assimilation over a 7 day period.

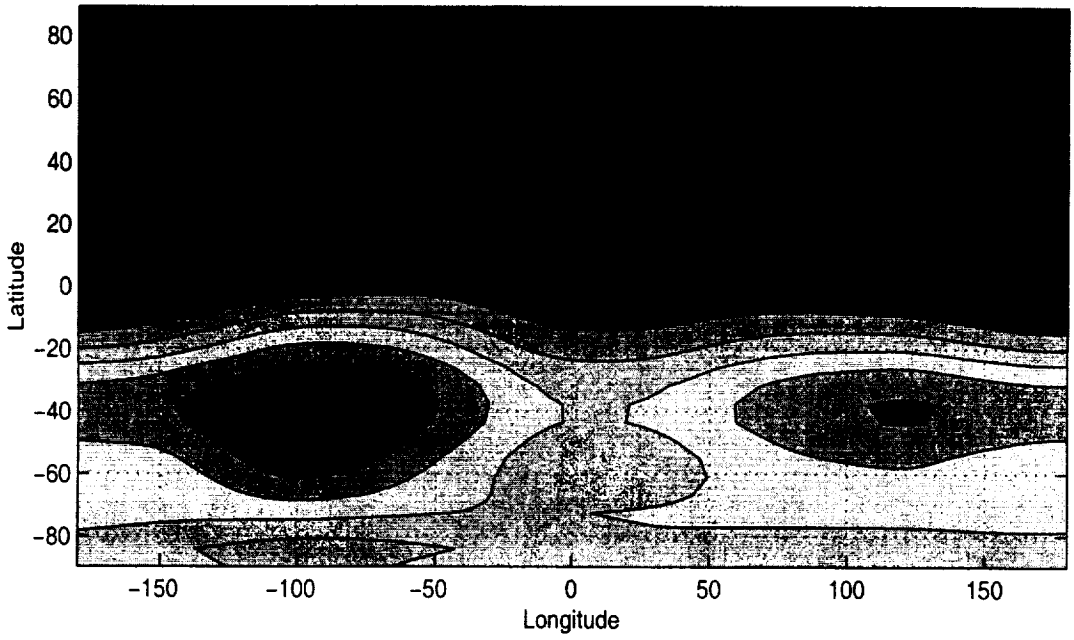
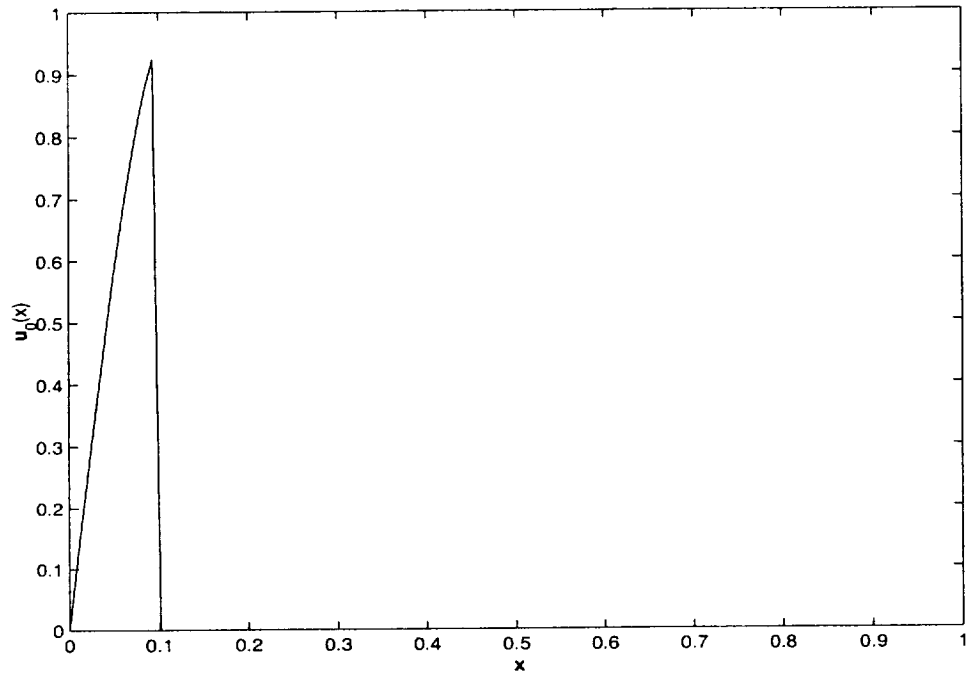
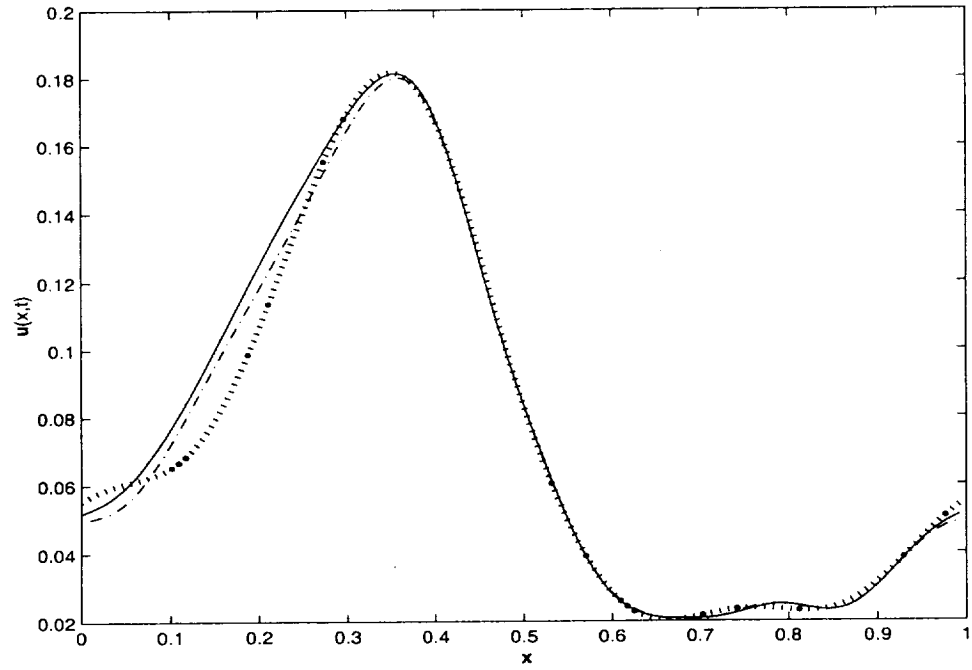


Figure 8: Reconstructions of correlations on day 7 using the *initially* 32 largest wavelets on day 1



(a)



(b)

Figure 9: Result of single set (*ie.* one ensemble member from each case) of experiments using 8,16 and 128 wavelet coefficients in the covariance approximation. (a) Initial value of u(x,t), (b) final value of u(x,t) for L=128 (solid), L=16 (dash-dot) and L=8 (dotted). The largest differences between the full and approximate systems occur where there are no observations, $0 \leq x \leq .38$.

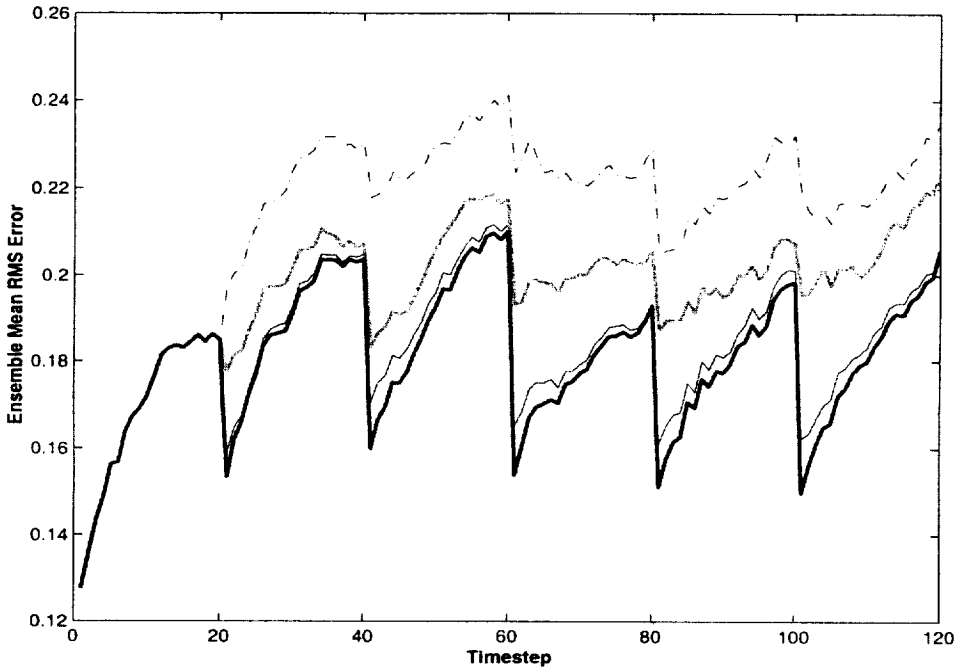


Figure 10: Ensemble mean RMS error for four assimilations experiments for the 1-d Burger's equation. L=128 (thick solid line), L=16 (thin solid line), L=8 (grey line) and L=4 (dash-dot). The 16 coefficient expansion experiment follows the full system closely for the entire time period, while the 8 and 4 coefficient experiments show progressively larger errors as the number of coefficients are reduced.

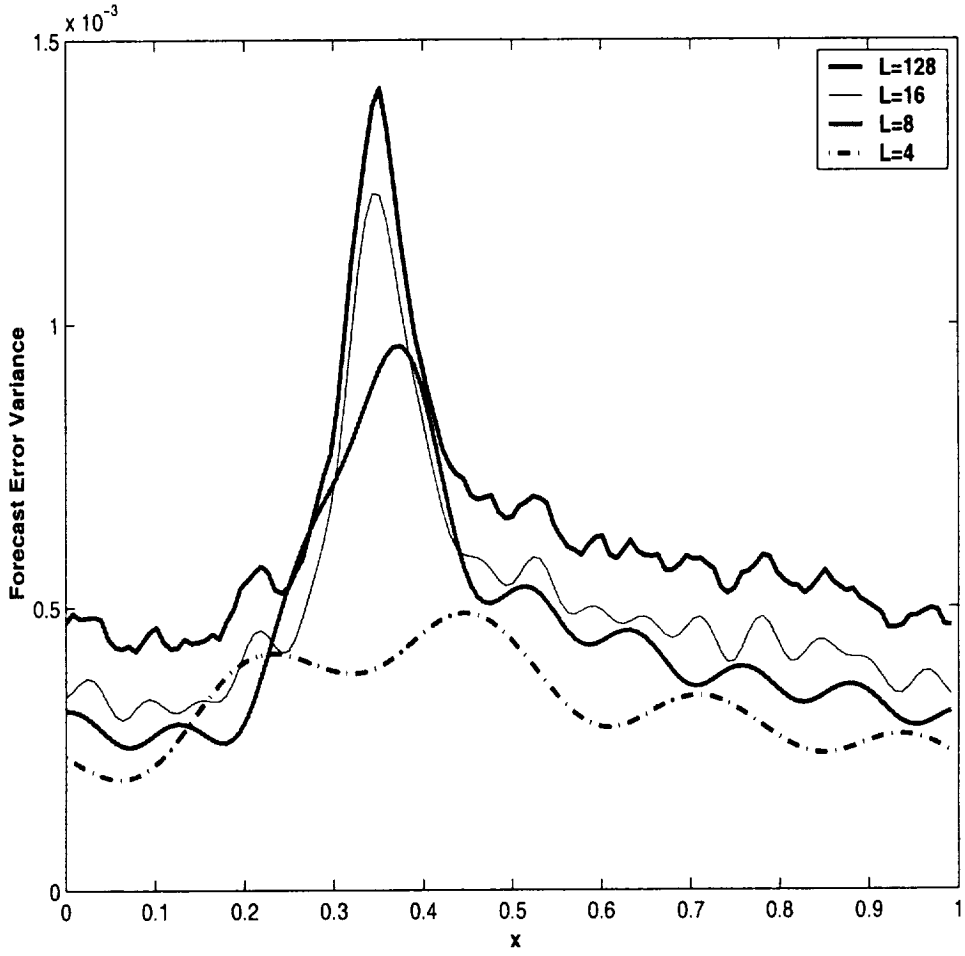


Figure 11: Diagonal from Pf for the with wavelet expansions of L=4 (dash-dot), L=8 (grey), L=16 (thin black) and L=128 (thick black).

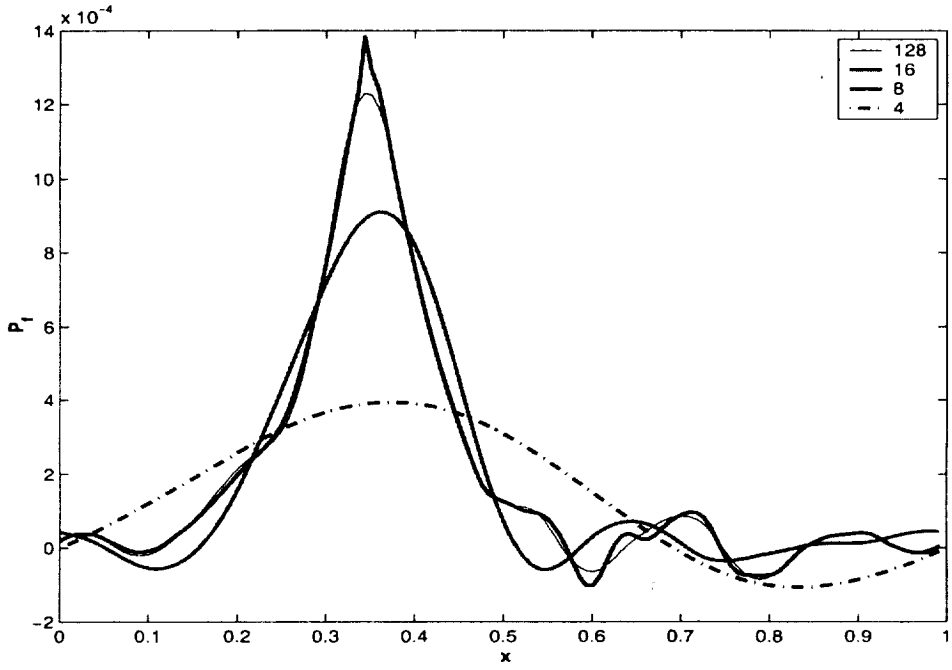


Figure 12: Covariance at point $x = 0.35$ at time $t = 120$. Except for some variance loss, the 8 and 16 term wavelet expansions accurately represent the covariance while the 4 term expansion misses most of the covariance structure.

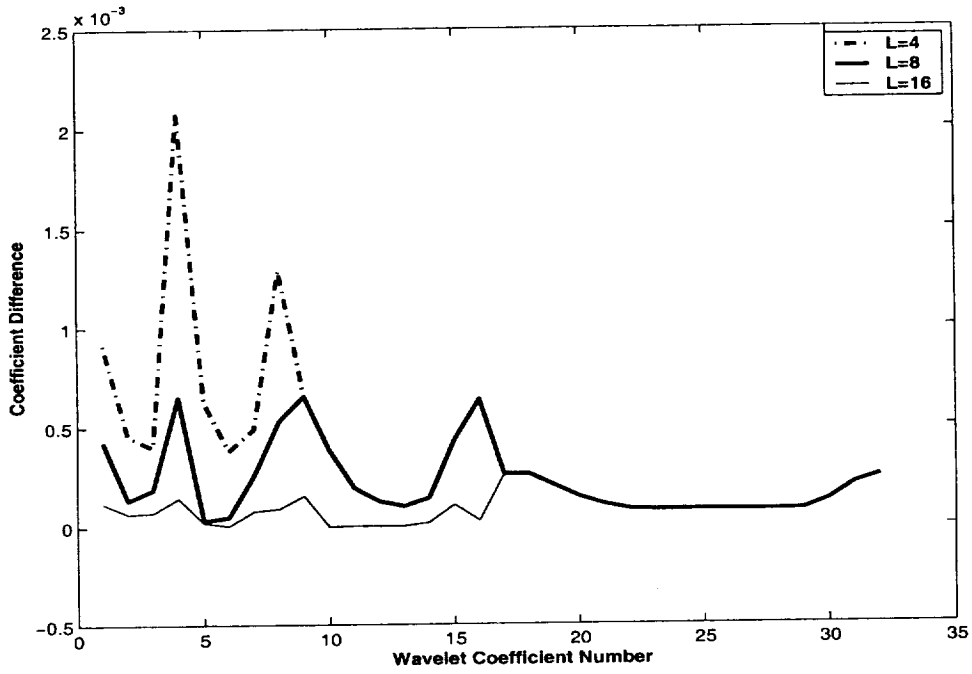


Figure 13: Difference between between full and partial covariance matrix diagonals in wavelet wavelet space. The Diagonal terms of $(\hat{P}_f)_{128} - (\hat{P}_f)_{16}$ is represented by the dashed line, while the solid line is $(\hat{Pf})_{128} - (\hat{P}_f)_8$.

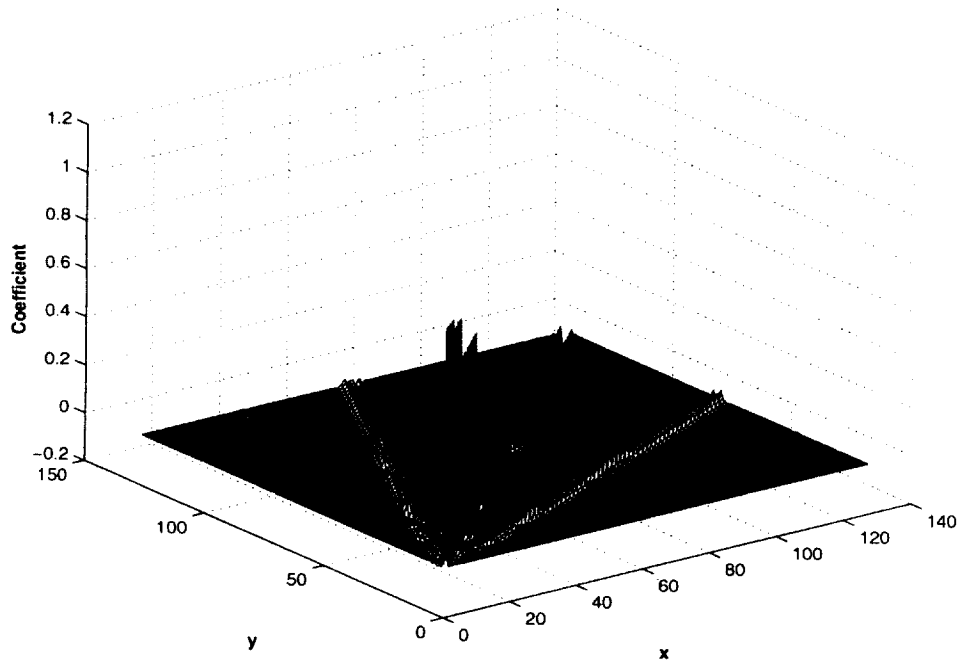


Figure 14: Projection of discretized tangent linear model onto wavelet basis. The wavelet representation of the dynamics is nearly diagonal and shows a exponential decay in diagonal elements with increasing scale j.